
EDUCATION**PEKING UNIVERSITY**

Beijing, China

Sep.2014 - Jul.2018

B.S School of Electronic Engineering and Computer Science, Computer Science Department

- Major in computer science, Major GPA: **3.6** / 4 Third year total GPA: **3.68**/4
- Core Curriculums: Introduction to machine learning, Biometrics, Calculus, Linear Algebra, Probability Statistics, Data structure & algorithms, Computer Networks, Operating Systems

UNIVERSITY OF CALIFORNIA, BERKELEY

Berkeley, USA

Aug.2018 - Jun.2019

M.Eng School of Electronic Engineering and Computer Science, Computer Science Department

- Major in computer science, Major GPA: **4.0** / 4

PUBLICATIONS

- [1] **Sheng Shen**, Daniel Fried, Jacob Andreas, Dan Klein: **Pragmatically Informative Text Generation**. (NAACL-HLT 2019)
- [2] David Bamman, Sejal Papat, **Sheng Shen**: **An Annotated Dataset of Literary Entities**. (NAACL-HLT 2019)
- [3] **Sheng Shen**, Yaliang Li, Nan Du, Xian Wu, Yusheng Xie, Shen Ge, Tao Yang, Kai Wang, Xingzheng Liang, Wei Fan: **On the Generation of Medical Question-Answer Pairs**. (arXiv – Submit to ACL 2019)
- [4] **Sheng Shen***, Zhenpeng Chen*, Ziniu Hu, Xuan Lu, Qiaozhu Mei, Xuanzhe Liu: **Ermes: Emoji-Powered Representation Learning for Cross-Lingual Sentiment Classification**. (WWW 2019, oral)
- [5] Zhenpeng Chen, Xuan Lu, **Sheng Shen**, Wei Ai, Xuanzhe Liu, Qiaozhu Mei: **Through a gender lens: An empirical study of emoji usage over large-scale android users**. (arXiv 2017 – WWW 2018)
- [6] **Sheng Shen**, Xuan Lu, Ziniu Hu, Xuanzhe Liu: **Towards Release Strategy Optimization for Apps in Google Play**. (Internetwork 2017)

RESEARCH EXPERIENCE*Berkeley, Nature Language Processing Group, BAIR***Project:** Leveraging Pragmatic Model for Text Generation

Aug. 2018-Now

Adviser: Dan Klein

- Devised a pragmatic model based reranking schema for the inference of summarization and data driven NLG tasks, which achieved general 1%-3% performance improvements over state-of-the-art results in terms of F1 score on all automatic metrics.
- Redesigned an incremental pragmatic model based decoding schema for the decoding process of summarization and data driven NLG tasks, which achieved an additional 1% performance improvement over the original pragmatic setting.

Results: Accepted by NAACL-HLT 2019*Tencent, Medical AI lab***Project:** On the Medical Question Answer Pair Generation

Apr. 2018-Sep. 2018

Adviser: Nan Du, Wei Fan

- Devised an entity-guided conditional variational auto-encoder (eg-CVAE) generator, which consider 3-pass decoding (type, entity and question), for medical questions generation
- Devised a unsupervised detector to ensure the relation between answer and crucial question split unchanged with retrieval and matching method on unstructured medical text books
- The whole framework improved the original QA system with 7% accuracy, and performs the best based on dialog generation metrics against baseline methods

Results: Submit a paper for ACL 2019 (first-author)*Peking University, Key Lab of High Confidence Software Technology, MoE***Project:** Ermes: Emoji-powered representation learning framework for cross-lingual sentiment classification Oct. 2017-Now**Adviser:** Xuanzhe Liu, Qiaozhu Mei

- Devised a distant learning task of predicting emoji to learn a general sentence-level representation, which aims to transfer sentiment information cross different languages
- Designed a hierarchical attention based Bi-LSTM/CNN network for distant learning phase, which achieved the accuracy by 0.45 (predict top 64 emoji)

• Redesigned a Multi-head attention based network for final document sentiment classification, which achieved the general accuracy around 0.81 and beat state-of-the-art method in nine tasks

Results: Accepted by WWW 2019 (co first-author, oral)

Project: Exploring Emoji Usage from Demographic Perspective

Oct. 2016-Jun. 2017

Adviser: Xuanzhe Liu, Qiaozhu Mei

- Defined, tested and selected active data set with all the demographic information from Kika
- Studied the correlation between the difference of emoji usage in gender, age and country and Human Development Index
- Applied SVM, Decision tree, GBDT based methods over sentence embedding to portray a user profile of gender and age based on the according usage of emoji

Results: Accepted by WWW 2018

Project: Towards Release Strategy Optimization for Apps in Google Play

Oct. 2016-Jul. 2017

Adviser: Xuanzhe Liu

- Collected the raw data and preprocessed the text information of apps in Google Play with tf-idf model, applied statistical method to detect combination effect of release with App ranking, rating trend and update purpose
- Purposed a more informative way to reveal release pattern as update intervals, applied general embedding and clustering method to detect local release pattern
- Generated a model to describe the update effects and the factors, which suggests the developer to improve their release strategy

Results: Accepted by Internetware 2017 (first-author)

University of Illinois at Urbana Champaign (UIUC), Interactive and Approximate Analytics Group

Project: Towards practical DATA-DIFF: Optimal Synthesis of Succinct Data Modification Scripts

Jul.2017-Sep.2017

Adviser: Aditya Parameswaran

- Formulate the practical problem, designed approximation algorithm to detect the non-overlapped operations user applied to update a database
- Purposed a simple policy gradient strategy based reinforcement learning model and a heuristic searching model to detect the addition and linear transformation operations used in an updated database
- Purposed a decision tree based clustering method to detect the operations applied over multiple attributes of a updated database

COMPETITION EXPERIENCE

Competition: CIKM Analytical Cup 2018 - Cross lingual short text matching

Jun. 2017-Aug. 2018

- Devised a graph based generator for augmenting the unbalanced labeled data
- Devised a multi-perspective context matching model for English and Spanish text matching

Rank: 8 / 1108

INDUSTRY EXPERIENCE

Kika Keyboard, Advertising Platform Bidding Group

Project: Advertising platform bid algorithm design and integration

Nov. 2017-Mar. 2018

- Participated in the design of a bidding algorithm based on logistic regression and Boosting strategy for different platforms
- Participate in advertising integration such as Conversant's direct customers, OpenX DSP's video ad integration, and traffic access Aerserver's technology integration

EXTRACURRICULAR ACTIVITY

Voluntary tutor at Bali island, Indonesia

Volunteer

Aug.2015

- Participated in the special English program teaching English to students

Debate Association, PKU

Leader of the debate team of EECS

Sep.2015 - Jul.2018

- Organized and taught members to debate with each other and registered for competition in campus

HONORS & AWARDS

- First prize in 2014 debate competition of PKU

Nov.2014

- Wusi Fellowship in PKU

Sep.2015, Sep 2017

- Second prize in the ACM competition in PKU Apr.2016
- Outstanding League Cadres Sep.2016
- Outstanding Student Researcher in PKU Sep.2017
- Distinguished Undergraduate Thesis in PKU Jul.2018